# 我国高校图书馆数据开放平台 FAIR 原则的应用评估分析

**张素芳[1] 梁睿 何国玉**

**华南师范大学 广州 510006**

**摘要：[目的/意义]**FAIR 原则是保障科学数据标识、溯源、共享及重用的数据管理原则，通过评估我国高校图书馆数据开放平台 FAIR 原则的应用情况，对于改善平台数据质量具有重要意义。**[方法/过程]**本研究通过网络调研，基于现有 FAIR 原则框架，构建适用于我国高校图书馆数据开放平台的 FAIR 评估框架，并从可发现性、可访问性、互操作性、可重用性四个维度对我国的高校图书馆数据开放平台进行评估分析。**[结果/结论]**基于以上研究，本研究从明确平台自身定位、提高数据质量、互通平台数据、规范数据标准、转化数据效能等方面提出高校图书馆数据开放平台建设改进建议。

**关键词:**高校图书馆；数据开放平台；FAIR 原则

# 1 引言 Introduction

习近平总书记在中共中央政治局第十一次集体学习时强调："发展新质生产力是推动高质量发展的内在要求和重要着力点""新质生产力已经在实践中形成并展示出对高质量发展的强劲推动力、支撑力"。2024 年 3 月 14 日，政府工作报告将"大力推进现代化产业体系建设，加快发展新质生产力"作为今年政府工作十大任务之首，并提出既要发展新质生产力，还要深入推进数字经济创新发展[1]。

数据要素是我国第五大生产要素[2]，作为数字经济时代的基础性和战略性资源，是新时代科技创新的关键要素。激活数据要素价值，推动生产力跃迁升级，

---

1 *本文系中国图书馆学会科研项目"基于知识生态链的中小型图书馆智慧化建设研究"（项目编号：2022LSCKYXM-FZZ-YB001）；广东省教育科学"十三五"规划 2020 年度研究项目"心理契约视角下粤港澳大湾区科研人员知识转移影响因素及促进机制研究"（项目编号：2020GXJK171）研究成果 。

作者简介：张素芳，副教授，博士，Email:342208681@qq.com;梁睿，本科；何国玉，硕士研究生

是在数字经济时代培育和发展新质生产力的核心要义[3]。

高校图书馆作为信息中心和数据信息资源中心，在新时代中肩负着教学科研，人才培养等重要任务，其数据开放是贯彻落实我国数据开放共享战略的重要环节[4]，如何规范化管理开放平台并推动数据要素向发展新质生产力流动是当前高校图书馆进行数据开放的核心任务。

# 2 文献综述 Literature Review

## 2.1 高校图书馆数据开放平台建设质量与要求分析

高校的使命在于人才培养和知识创新，是实施创新驱动发展战略、建设创新型国家的重要力量[5]。高校图书馆作为重要的教辅部门，具有教育和信息服务两大主要职能，提供学科建设、人才培养、科学研究等服务[6]，以协助高校实现服务国家战略、支持民族复兴的历史使命[7]。其中，科研数据管理成为高校图书馆实现自身价值的重要途径[4]。国内多所一流高校先后开展了科研数据管理服务工作，推进相关管理措施和政策的制定[8]。随着开放数据运动的深入以及科研数据管理需求的提升，高校图书馆需要进一步重视数据开放平台的建设质量，以标准化、科学化的方式存储数据，从而提高科研数据的质量、扩大科研数据的共享范围[9]。然而就实践情况来看，数据开放平台仍存在着诸多问题，如数据难以质量保障、用户访问困难[10]、共享水平较低、数据安全保障体系薄弱[11]等，因而，数据开放平台建设质量的评估研究引起学者的关注，数据的开放共享需要科学的数据管理原则予以指导。

## 2.2 FAIR 原则

FAIR 原则是“代表学术界、工业界、资助机构和学术出版商的各种利益相关者聚集在一起，设计并共同批准的一套简明且可衡量的原则”[12]，是可发现(Findable)、可访问(Accessible)、可互操作(Interoperable)、可重用(Reusable)四个维度的首字母组合，用以保障科学数据标识、溯源、共享及重用。

### 2.2.1 FAIR 原则定义相关研究

自 FAIR 原则提出以来，各国政府、国际组织以及学术团体等就展开了对其概念内涵和具体指向的一系列的研究，不断细化对该原则的理解和应用要点。2018 年 11 月，美国地球物理联合会（AGU）数据项目总监 Shelly Stall 做了“赋能 FAIR 数据：地球、太空与环境科学”的报告[13]，对 FAIR 原则的内容、参与机构、利益相关者及其责任等相关内容进行阐述。2020 年，Annika Jacobsen 基于不同领域对 FAIR 原则的理解歧义，提出要简明扼要地界定 FAIR 原则并为一些创新领域留下余地[14]。

### 2.2.2 FAIR 化的相关研究

随着 FAIR 原则的研究逐步深入，越来越多学科领域的组织、机构注意到 FAIR 原则，并将其广泛应用于各大学科领域的实践活动中。2018 年 11 月 7 日，在第二届欧洲开放科学云（EOSC）峰会上，FAIR Data Expert Group 发布了 Turning FAIR Data into Reality[15]和 FAIR Data Action Plan[16]两份研究报告，进一步推进全球数据 FAIR 化；2019 年 10 月，研究数据联盟（RDA）全体会议组织了“勾画未来的数据蓝图：FAIR 语义与 FAIR 存储库”研讨会，针对存储库与 FAIR 融合、语义 FAIR 化、FAIR 产品设计等问题展开讨论[17]；政府间组织 ELIXIR 开展了“农作物大规模基因分型数据集”的 FAIR 化（FAIR-ification）研究工作，以确保分布式机构存储库中的表型数据的互操作连接，为作物育种提供数据支撑[18]。

## 2.3 FAIR 原则评估研究

### 2.3.1 FAIR 评估框架概述

FAIR 原则被广泛应用到各大学科领域的实践活动后，出现了对于 FAIR 原则的歧义和操作上的不一致，背离了 FAIR 原则推动共享的初衷。同时，FAIR 原则实施的效果如何、有待改进的问题在哪里，也需要进行调查。因而，FAIR 原则应用评估应运而生。Peter Wittenburg 等[19]基于 Wikinson 构建的指标体系评估了欧洲研究基础设施战略论坛（ESFRI）项目的 FAIR 化程度，同时解释了进

行 FAIR 化评估的注意事项。德国学者 Lars Quakulinski 甚至建立了一个适用于不同领域的，同时又低门槛和定义明确的人为错误研究框架，旨在促进在各领域相关者之间有效内容传播[20]。

但是，在实际活动中应用 FAIR 原则存在着评估指标不统一，评估手段困难等挑战，为此为了有效评估数据资源的 FAIR 实施情况，为开放数据管理提供明确的指导和标准，相关组织展开了 FAIR 评估框架的研究工作，其中就包括评估指标、评估工具与评估方法的建设与应用。

目前国际上通用的 FAIR 原则评估框架根据适用对象分可以分为两类，第一类是适用于多种不同学科领域的通用框架，比如 Wilkinson 等于 2018 年 7 月最早提出的 FAIR 评估模型——FAIR-ness Maturity Indicators(MI) 以及 Go FAIR Metric Group (GFMG)的 FAIR 通用指标框架[21]; 第二类是针对某一学科领域数据的专用评估框架，通常是结合第一类框架以及调研对象的数据特点设计的，如李春秋等根据国外已有的 FAIR 应用评估框架，结合我国医学科学数据开放平台的特征，设计出面对医学科学数据的 FAIR 应用评估指标体系，用于评估我国 11 个样本平台的 FAIR 原则应用情况[22]；爱达荷大学的学者 Alowairdhi 利用 FAIR 原则设计出针对 Earth Science Data Management and Stewardship 的可实施框架，用以来改善地球科学数字资源的数据管理和管理[23]。文化领域，Lukas Koster 为实现不同 LAM（libraries, Archives, Museums）机构馆藏数据的可访问和可重用，解决不同结构馆藏数据重用、许可限制等问题，结合 LAM 机构馆藏数据的特征，提出了一系列针对 LAM 的 FAIR 原则评估指标[24]。

根据评估工作的可操作性，FAIR 原则评估框架可以分为多级框架和指南性框架。前者是进一步将评估指标细化到不同层级，方便对评估结果分级，如荷兰数据存档与网络服务(Dutch Data Archiving and Networked Services, DANS)[25]、欧盟 Horizon 2020 及澳大利亚研究数据共享组织(Australian Research Data Commons, ARDC)的 FAIR 原则评估指标体系[26]、RDA FAIR 评估工作组 2020 年上半年推出的 FAIR 数据成熟度模型（FAIR Data Maturity Model，FDMM）；后者有详细的指南，便于实际评估实施，比如澳大利亚联邦科学与工业研究组织(Commonwealth Scientific and Industrial Research Organisation, CSIRO)的 5 星级数据评估工具[27]以及 FAIRsFAIR 工作组的评估框架[28]。

### 2.3.2 FAIR 原则评估流程

FAIR 原则的评估流程可根据实际情况大致分为两类，一类是人工评估，另一种则是利用各种 FAIR 评估软件进行的自动评估。关于人工评估，Annika Jacobsen 等描述了一个通用的分步 FAIRification 工作流程[29]，适用于任何类型的数据，在“Bring Your Own Data”（BYOD） 研讨会上得以应用，分为三阶段六步骤：（1）前 FAIR 化阶段，确定目标即确认评估对象，分析（元）数据；（2）FAIR 化阶段 ，定义（元）数据的语义模型、使（元）数据可链接、存放 FAIR 数据；（3）后 FAIR 化阶段，评估 FAIR 原则应用情况。自动评估则是针对大部分机器可读数据的处理过程，包括数据收集、数据准备、数据分析、预测建模和模拟等步骤，将 FAIR 原则融入到计算工作流中，促进了对数据质量的评估，并产生了新的数据，提高了数据的标准化和可重用[30]。

CESSDA（欧洲社会科学数据档案委员会）参照欧盟的 FAIR 实施框架进行了组织科学数据的自我评估[31]，由此发现在 FAIR 文化和生态建设上的不足;在环境科学领域， WDCC（世界气候数据中心）采取一系列 FAIR 原则的应用措施[32]，旨在为更好地对接地球观测、气象、海洋学和环境科学的数据中心。

# 3 我国高校图书馆数据开放平台 FAIR 应用评估指标体系的构建　Construction of FAIR application evaluation index system of academic libraries’ open data repositories in China

## 3.1 框架对比分析

本研究采用人工评估的评估流程，首先选取国内高校图书馆数据开放平台作为对象，分析平台内（元）数据，经过综合评估，优先选择了适用学科范围较广的评估框架（GO FAIR）来作为基底，然后再从指标设计的角度出发，选择便于评估指标重要程度以及分析 FAIR 化程度的评估框架（RDA、ARDC），同时为考

虑实际评估工作开展的顺利进行，选择可实施性更强的评估框架（FAIRsFAIR）作为补充。至于指标体系的专业适用性，因为高校图书馆数据共享平台类型多样、数据庞杂，但是并不在特定领域深耕，所以结合了 Lukas Koster 针对 LAM（图书馆、艺术馆、博物馆）特点的 FAIR 原则应用研究[24]，构建出适用于评估我国高校图书馆数据开放平台的评估框架体系。

有关上述 4 种框架的比较分析、选取原因以及缺点如表 3-1 所示。总体而言，4 个评估指标框架各具特色，也都有缺点，本文通过不断地进行调整和改进，择优去劣，从而选择出适用于评估高校图书馆数据开放平台的评估指标。

表 3-1 FAIR 评估框架对比

Table 3-1FAIR evaluation framework comparison

| 名称 | 评估重点 | 优点 | 缺点 |
| --- | --- | --- | --- |
| GO FAIR | 通过半自动化过程进行评估，强调数字对象的机器可读性 | 明确指标适用原则、评估原因、评估方法和适用范围 | 评估内容宽泛，模糊性强，评估结果难确定 |
| RDA | 重点评估元数据描述性、标识符采用情况 | 划分指标优先级，将指标分为必要指标、重要指标和一般指标，揭示评估指标的重要程度 | 指标设置较多，评估流程较为繁琐 |
| FAIRsFAIR | 划分指标优先级，将指标分为必要指标、重要指标和一般指标，揭示评估指标的重要程度 | 为每条细则制定一至多个评估指标，并设置分层模型进行测试，评估可实施性强 | 未给出具体的评估方法，指标测试有一定难度 |
| ARDC | 采取赋分制形式问卷评估，对各项指标的 FAIR 化程度进行严格划分 | 指标简洁明了，以问卷形式呈现，容易知道数据 FAIR 化程度 | 可重用维度只有 1 个指标 |

## 3.2 评估指标确定

确定指标分两步走，首先参考上述四个评估框架从原框架中抽取适用于评估高校图书馆数据开放平台的评估指标，形成评估框架原型；然后再结合图书馆数据开放平台的数据特征以及 Lukas Koster 针对 LAM 的 FAIR 原则评估框架，对评估框架原型的各项指标进行微调，优化。

评估框架原型具体内容如表 3-2 所示。

表 3-2 高校图书馆数据开放平台评估框架原型

Table 3-2 University library open data repositories evaluation framework

prototype

| 一级指标 | 二级指标 | 指标借鉴情况 |
| --- | --- | --- |
| 可发现性（Findable） | F1.1 为数据分配全局唯一标识符 | GO FAIR、RDA、FAIRsFAIR、ARDC |
| | F1.2 为数据分配持久标识符 | GO FAIR、RDA、FAIRsFAIR、ARDC |
| | F2 元数据包括描述性核心元素 | RDA、FAIRsFAIR |
| | F3 元数据包括它所描述的数据的标识符 | RDA、FAIRsFAIR |
| | | RDA、FAIRsFAIR |
| | F4（元）数据在可搜索资源中注册或索引 | RDA、FAIRsFAIR |
| 可访问性（Accessible） | A1 可通过标准化通信协议访问数据 | RDA、FAIRsFAIR、GO FAIR |
| | A2 元数据包含数据的访问级别和访问条件 | GO FAIR、ARDC、FAIRsFAIR |
| | A3 即使数据不再可用，也可以访问元数据 | GO FAIR、RDA |
| 可互操作性（Interoperable） | I1 元数据使用形式化知识表示语言表示 | RDA、ARDC、FAIRsFAIR |
| | I2 元数据包括数据与其相关实体之间的链接 | |
| | I3 数据以什么（文件）格式提供？ | ARDC |
| 可重用性（Reusable） | R1 元数据包括可以重复使用数据的许可证信息 | RDA、ARDC、FAIRsFAIR |
| | R2（元）数据与详细出处相关联 | GO FAIR、FAIRsFAIR、ARDC |

对上述指标进行取舍的考虑如下：在可发现性维度下，大部分指标均适用，但考虑到数据标识符持久性以及唯一性评估的手段及指标的不同，故将 F1：“数据被分配一个全局唯一和持久的标识符”拆分成两个小维度（F1.1、F1.2），以便于实际工作的开展与后续结果分析；在可访问性维度中的各项指标均适用于实际评估，仅将 GO FAIR 中的 A1 下的两个小维度维度拆分成两个维度(现 A1、A2），原因是元数据协议与元数据的访问限制两个维度在图书馆数据开放平台的评估的手段及指标体现是不同；在可互操作性维度，只剔除了 I2：“（元）数据使用遵循 FAIR 原则的词汇表”这一指标，原因是在大多数数据集，FAIR 原则词汇表即元数据描述标准，与 F2 指标重合，无需重新评估；在可重用性维度没有需要修改或剔除的指标。同时还剔除了部分不具备评估条件或不适用于调查高校图书馆数据开放平台的指标，如删除了 RDA 评估框架中“元数据可以手动访问”(RDA-A1-02M)和“数据可以手动访问”(RDA-A1-02D)等由于实际情况难以评估的指标以及图书馆数据开放平台中不存在的指标如有关社区标准的四条指标

（RDA-R1.3）。

最终得到针对高校图书馆数据开放平台的 FAIR 评估框架，如表 3-3 所示。

表 3-3 高校图书馆数据开放平台评估框架

Table 3-3 Evaluation framework of university library open data repositories

| 一级指标 | 二级指标 | 三级指标 |
|---|---|---|
| 可发现性 (Findable) | F1 标识符 | F1.1 标识符的唯一性 |
| | | F1.2 标识符的永久性 |
| | | F1.3 元数据包含其所描述数据集的标识符 |
| | F2 元数据 | F2.1 元数据标准规范 |
| | | F2.2 元数据丰富度 |
| | F3 搜索引擎发现性 | F3.1 在平台中的登记注册情况 |
| 可访问性 (Accessible) | A1 访问条款 | A1.1(元)数据访问条款 |
| | A2 访问协议 | A2.1 (元)数据访问协议 |
| | | A2.2 (元)数据标访问协议的标准性 |
| | A3 访问权限 | A3.1 用户审核机制 |
| | A4(元)数据寿命方案 | A4.1(元)数据存储机制 |
| 可互操作性 (Interoperable) | I1 数据引用 | I1.1 合规的数据引用方式 |
| | I2 数据关联 | I2.1 关联到相关数据集 |
| | | I2.2 关联到相关数据（外部数据、相关论文） |
| | I3 数据文件格式 | I3.1 使用通用的机器可读的格式表示 |
| 可重用性 (Reusable) | R1 数据重用许可 | R1.1 数据集使用许可声明 |
| | | R1.2 说明数据集的限制原因 |
| | R2 数据溯源 | R2.1 数据溯源信息 |
| | | R2.2 溯源信息格式的标准性 |

该框架以 FAIR 原则的四大维度延展，在可发现、可访问、可互操作、可重用 4 个一级指标下归纳了 12 个二级指标，划分了 19 个三级指标。其中，从实际需求出发，二级指标根据图书馆数据内容以及类型，从评估框架原型中的细分指标中进行归纳，与原指标框架的顺序存在细微差异。而部分三级指标是在不影响原有含义的基础上，对指标进行微调或修改，使其指标内容更容易理解，更适用于评估高校图书馆数据开放平台。

在可发现性维度中，考虑到图书馆数据开放平台有的数据集尽管规范不标准，但其元数据也十分丰富，故不能将其视为一个指标来评估，因而归纳出 F1 标识

符、F2 元数据、F3 搜索引擎可发现性三个二级指标。将评估框架原型中的 F2 拆分成 F2 中的两个三级指标（F2.1、F2.2），；LAM 框架中涉及到的 FAIRsharing 和 Re3data 是国际权威的数据仓储目录，在其中注册并通过认证的数据平台和存储库高度开放且更加标准化，故将原 F4：“（元）数据在可搜索资源中注册或索引”这一指标的评估细则定为有无在 FAIRsharing 和 Re3data 注册。

可访问性维度，归纳出 A1 访问条款、A2 访问协议、A3 访问权限、A4(元)数据寿命方案四个二级指标。首先在原 A1：“可通过标准化通信协议访问数据”细分出 A1“访问条款”这一指标，以区别访问数据集与使用数据集；另外与数据相关的所有人员都参与到数据集的管理中去，可建立一个咨询小组，用于沟通使用要求以及用例概述等，这也是下文多个平台用户组机制的由来，故设置 A3.1：“用户审核机制”这一指标。荷兰数字遗产和文化知识中心 DEN 以荷兰语“DE BASIS”发布了一套文化遗产数字化的最低要求[33]，包括权利管理、可查找性、创建、展示、数据可持续性/保存、描述，其中便提及了数字馆藏数据的数据可持续性/长期保存性的重要程度，为此将原 A3 细化到 A4.1：“(元)数据存储机制”，以便检查平台数据的长期保存性。

在可互操作性维度，归纳出 I1 数据引用、I2 数据关联、I3 数据文件格式。将原指标 I1：“元数据使用形式化知识表示语言表示”修改为现指标 I1：“数据引用”，以便于下面了 I3：“数据文件格式”重合；指标 I3：“数据文件格式”重点评估数据资源的格式是否机器可读，本文把数据文件格式分成了上传格式和下载格式，其原因在于由于尽管有的数据开放平台规定了上传的数据文件格式，但用户能下载到的数据文件格式与其声明的数据文件格式并不一致。

在可重用性维度，归纳出 R1 数据重用许可、R2 数据溯源两个二级指标。为了避免在每种情况下反复请求重用许可的必要性，应立即明确允许对对象和元数据执行哪些操作，这一工作可通过为数据集提供可重用许可声明，比较常用的许可声明有 the Creative Commons (CC) license，最常用的版本就是 CC-BY 和 CC-0，故将评估细则定为 CC-BY、CC-0 等国际通用标准许可；同时对 R1、R2 指标进行划分，便于评估与分析。

除去上述各维度的指标调整过程，另外还对各框架的部分指标进行了取舍，比如 RDA 评估框架中将对象分成数据与元数据，但 LAM 机构的数据对象分为

对象（如书籍、期刊、工件、视频、数据集等）、元数据（例如标题、创建者、标识符、日期等）和元数据记录（有关特定数据库中对象的元数据元素的正文）三种，因此本文的调研对象类型应该为元数据记录。LAM 框架分别对三种对象设计了不同的适用指标，根据适用指标内容，将元数据与数据两者进行了合并，如“元数据可通过免费协议访问”(RDA-A1.1-01M)、“数据可通过免费协议访问”(RDA-A1.1-01D)两条指标都是针对评估对象采用的访问协议及其性质进行评估，核心评估内容一致，将其合为 A2.2：“(元)数据标访问协议的标准性”。

# 4 我国高校图书馆数据开放平台 FAIR 应用现状分析

# Analysis on the current situation of FAIR application in open data repositories of academic libraries in China

## 4.1 样本选择

国内高校图书馆数据开放平台的实践情况还不够成熟，且由于建设经验的缺乏，当前国内高校图书馆很少能够独立开展相对完整的开放数据服务，更多的是高校图书馆或对应学科科研部门独立或合作开展服务[10]。因此本研究认为在调研国内高校图书馆开放数据平台时视角可以适当放宽，高校内部机构与高校图书馆开展的开放数据平台均可以纳入调研范围。因而本研究所调研的高校图书馆数据开放平台主要分为两种类别：一类是高校图书馆与校内其他机构增强联系紧密合作；一类是由高校中某个机构或部门独立开展。

确定选择条件后，本研究在 bing 搜索引擎、各大高校官方网站上进行检索，最终检索到符合条件的我国高校数据开放平台共 10 个，但是由于访问权限的限制，只有四个平台可以访问。另外在选取调研对象的期间，除了上述由高校本身开发的开放数据平台，还有部分高校选择提供第三方的开放数据平台，如中国传媒大学、清华大学、浙江财经大学等高校，均在本校图书馆提供 CnOpenData 开放数据平台的入口，因此，本文将 CnOpenData 也列为调研对象。故而本文最终

的调研对象共五个，分别为北京大学开放研究数据平台、浙江大学开放数据平台、复旦大学社会科学数据平台和华东师范大学人文社科大数据平台（以下简称为“北京大学”、“浙江大学”、“复旦大学”、“华东师范大学”）以及 CnOpenData。

## 4.2 可发现性

数据的可发现性是影响数据发挥其价值的重要因素，数据只有被用户发现，才有被使用、分析、组织的可能。支撑数据可发现性的三个重要指标特征包括唯一永久性标识符、元数据丰富程度以及搜索引擎可发现性。可发现性的各指标评估结果如表 4-1 所示。

表 4-1 可发现性维度各指标评估结果

Table 4-1Evaluation results of finable dimension indicators

| 平台 | 北京大学 | 复旦大学 | 华东师范大学 | 浙江大学 | CnOpenData |
|---|---|---|---|---|---|
| 标识符 | | | | | |
| 是否声明标识符 | √ | √ | √ | × | √ |
| 永久性标识符 | DOI | HDL | HDL | 否 | 否 |
| 本地标识符 | × | × | × | × | √ |
| 标识符类型 | 永久性 Web 标识符 | 永久性 Web 标识符 | 永久性 Web 标识符 | 无标识符 | 本地标识符 |
| 元数据 | | | | | |
| 元数据标准声明 | √ | √ | √ | × | × |
| 遵循的元数据标准 | DDI、DataCite 等 | DDI、DataCite 等 | DDI | - | - |
| 元数据标准类型 | 国际通用元数据标准 | 国际通用元数据标准 | 国际通用元数据标准 | 未声明元数据标准 | 未声明元数据标准 |
| 元数据格式丰富度赋值 | 复用国际元数据标准格式 | 复用国际元数据标准格式 | 复用国际元数据标准格式 | 自定义非标准格式 | 仅有标题和非标准简单描述 |
| 搜索引擎 | | | | | |
| Re3data 注册情况 | √ | √ | √ | × | × |
| FAIRsharing 注册情况 | × | × | × | × | × |
| 本地索引 | √ | √ | √ | √ | √ |

#### 4.2.1 标识符

标识符 PID 是标识数据身份的一系列字符，通过检索协议将数据与资源进行关联，标识符的永久性是确保图书馆数据被发现的关键，适用于图书馆数据的 PID 有 Handle (handle.net), DOI (doi.org), URN-NBN 和 URL 等，但是 URL 这种类型的内部标识符只推荐使用与基于特定系统或域的 Web 上，但它不是全局和持久标识符。

国内外各高校平台调查结果如表所示，5 个样本平台有 4 个平台为数据集声明了标识符，其中有 3 个平台使用永久性标识符，而 CnOpenData 虽然声明采用了标识符，但并未明确指出其标识符类型与名称，而浙江大学并未声明其采用了标识符，只在本地提供数据集的 URL。

#### 4.2.2 元数据

在通过元数据描述各类资源的过程中，基于不同资源的不同特点，按照整合和共享的要求，需要制定通用的元数据标准，以科学、准确、全面的描述资源的属性和特征，目前国际通用的元数据标准有都柏林核心(DC)、DCAT、Schema.org、CKAN 等，国内外各数据开放平台调查结果如表所示，有多个平台通过复用元数据标准构建元数据方案，元数据较丰富。国内平台遵循的标准较国外平台的要缺乏规范性和标准性，北京大学和复旦大学都是基于哈佛大学开源软件 Dataverse 搭建测试数据平台，故其规范性也较好，因而元数据格式较为校准，有着多种元数据标准形式，如 DDI、DataCite、Dublin Core 等，另外元数据丰富度也较为完善；浙江大学和 CnOpenData 虽提供元数据格式但未声明遵循的元数据标准，不过都有较完整且丰富元数据元素，但是 CnOpenData 仅展示数据集元数据的各字段及其内容，而浙江大学除了数据集的元数据字段内容，还有元数据学科分类、负责人、发布日期等相关信息。

#### 4.2.3 搜索引擎

re3data 是为了应对在数据共享增长背景下,越来越多的不同类型的科研数据知识库 (RDR)的诞生而导致 RDR 的异构性而制定的，其中的 re3 分别代表

registry、research、repositories 这三个单词，其全称是 Registry of Research Data Repositories，即科学数据仓储注册系统，其旨在对所有领域的科学数据仓储基于注册机制进行索引化和结构化描述，采用信息图标来描述每个科学数据仓储的基本特征，以便使用者能快速定位要寻找的 RDR。同时 FAIRsharing 也是国际权威的数据仓储目录，它提供了关于数据标准、数据库、存储库和政策的信息，以及与其他支持 FAIR 的资源互操作的搜索和可视化工具和服务，有利于数据使用者发现、选择和使用符合规范标准的数据库、存储库，并使数据上传者使的资源更容易被发现、更广泛地采用。在 Re3data 和 FAIRsharing 注册并通过认证的数据平台和存储库高度开放且更加标准化。

5 个样本平台中有 3 个平台仅在 Re3data 注册，有 2 个平台既没有在 Re3data 注册，也没有在 FAIRing 注册。5 个样本平台均有在本地标引数据集并提供多样化的检索方式，比如北京大学，复旦大学，和华东师范大学提供高级检索；；CnOpenDataData 虽不提供高级检索，但是可以具体定位到数据集某个具体部分的数据且收录数据较完整；浙江大学在这方面则有所欠缺，仅支持对检索结果按学科、关键词、发布日期和负责人分类进行筛选数据。

## 4.3 可访问性

当用户需要获取数据时，他们会考虑如何访问这些数据。为了保证数据的可访问性，需要在遵守访问协议的前提下，确保用户能够轻松地获取（元）数据。值得注意的是，可访问性并不意味着所有数据都必须公开，而是根据数据的性质确定公开的内容和时间。可访问性的各指标评估结果如表 4-2 所示。

表 4-2 可访问性维度各指标评估结果

Table 4-2 Evaluation results of accessible dimension indicators

| 平台 | 北京大学 | 复旦大学 | 华东师范大学 | 浙江大学 | CnOpenData |
|---|---|---|---|---|---|
| 访问条款 | | | | | |
| 用户条款声明明确性 | 有声明用户条款 | 有声明用户条款 | 有声明用户条款 | 有声明用户条款 | 无声明用户条款 |
| 用户条款声明标准性 | 规范、详细的用户条款 | 不规范、简略的用户条款 | 不规范、简略的用户条款 | 不规范、简略的用户条款 | 无声明用户条款 |

| | | | | | |
|---|---|---|---|---|---|
| 访问协议 | | | | | |
| 访问协议声明 | √ | √ | √ | √ | √ |
| 访问协议类型 | HTTP、API | HTTP | HTTP | HTTP | HTTP |
| 在线访问与获取 | √ | √ | √ | √ | √ |
| 访问权限 | | | | | |
| 平台审核机制声明 | √ | √ | √ | √ | √ |
| 平台审核机制内容 | 用户注册后部分开放数据需申请 | 用户注册后部分开放数据需申请 | 用户注册后部分开放数据需申请 | 用户注册后部分开放数据需申请 | 用户注册后部分开放数据需申请 |
| 用户审核机制声明 | √ | √ | √ | √ | × |
| （元）数据存储方式 | | | | | |
| 存储平台 | DataVerse | DataVerse | 自身平台 | 自身平台 | - |

### 4.3.1 访问条款

5 个样本平台中有 4 个平台声明了用户访问条款，其中北京大学和复旦大学均是在平台首页处有声明用户使用条款声明，虽然均是 Dataverse 搭建的测试数据平台，但是北京大学在用户条款中的声明较为详细且全面，复旦大学仅对用户隐私信息和服务条款等方面做了声明，对数据集使用方面的声明并未提及；华东师范大学只在具体数据集元数据部分处有用户条款声明，且用户条款内容均是关于数据集的使用与下载方面内容；浙江大学虽在具体数据集元数据部分中有条款的页面，但是只有“访问权限”和“版权及所有权声明”两个部分，且内容较为简略；CnOpenData 无论在平台首页还是具体数据集中都无法找到用户条款界面。

### 4.3.2 访问协议

5 个样本平台均支持 HTTP 协议访问和数据下载。从这个方面来说，FAIR 所要求的标准化访问协议环境已经得到了满足，但除此之外，北京大学还提供 API 接口。

### 4.3.3 访问权限

访问权限既包括平台对用户的审核，也包括上传者对用户的审核，也就是用户审核机制。

首先是平台对用户的审核。北京大学、复旦大学、华东师范大学、浙江大学对于完全开放的数据集，未注册登录的用户都可以访问并下载，但是对于有权限要求的数据集，则需要注册并登录才可访问，甚至还需进行申请访问之后才可以进行访问并下载；在 CnOpenData 平台，用户虽然无需注册便可访问并浏览数据集及其内容，但是需注册并登陆才可下载数据集，同时部分数据需要开通会员才可访问并下载。

在上传者对用户的审核方面，CnOpenData 作为一个开放数据整合平台，并不提供用户上传数据服务，故不做讨论。其余 4 个平台均提供用户审核机制，各个平台间提供不同权限层次的数据共享审核机制，比如北京大学将用户分为 3 个非平行的基本权限组，数据集和数据空间中的权限角色都有细分，较为详细全面；华东师范大学中有四种权限，分为成员、编辑、管理员和访问未发布的数据集；浙江大学平台提供默认角色组以及权限，但同时也支持自定义角色组和权限分配操作，更加人性化。

注:

### 4.3.4 （元）数据存储方式

随着时间的推移，数据集往往会消失或失去利用价值，即使数据不可再用，元数据也可以访问，因此元数据应保存到可靠、稳定且专业的存储平台中，并且提供元数据保存声明。由于 CnOpenData 并不提供用户上传数据服务，故不做讨论，其余 4 个样本平台均有声明数据保管措施并承诺提供稳定存储，存储策略各有差异。北京大学和复旦大学基于 DataVerse 存储库软件保存，而华东师范大学和浙江均基于自身研发平台保存。

## 4.4 可互操作性

可操作性指让机器在访问、关联、集成不同来源的数据时，能够更加准确、顺畅地理解，从而为用户方便获取数据奠定基础。此外，可操作性还强调人类和机器对数据的交互与理解，以便更好地实现数据的利用和重用。可互操作性的各指标评估结果如表 4-3 所示。

表 4-3 可互操作性维度各指标评估结果

Table 4-3 Evaluation results of interoperable dimension indicators

| 平台 | 北京大学 | 复旦大学 | 华东师范大学 | 浙江大学 | CnOpenData |
| --- | --- | --- | --- | --- | --- |
| 数据引用 | | | | | |
| 是否提供数据引用方式 | 提供且格式丰富 | 提供但格式单一 | 提供但格式单一 | 提供但格式单一 | 不提供 |
| 数据关联 | | | | | |
| 是否有数据关联 | √ | √ | √ | √ | √ |
| 数据关联内容 | 提供但格式单一 | 提供但格式单一 | 提供且格式丰富 | 提供但格式单一 | 提供且格式丰富 |
| 文件格式 | | | | | |
| 上传数据文件格式 | 非机器可读的数据文件格式 | 数据文件格式不做要求 | 机器可读的数据文件格式 | 数据文件格式不做要求 | 不能上传文件 |
| 下载数据文件格式 | 采用结构化、开放标准、机器可读的格式 | 采用结构化、开放标准、机器可读的格式 | 采用结构化、开放标准、机器可读的格式 | 采用结构化、开放标准、非机器可读的格式 | 采用结构化、开放标准、非机器可读的格式 |

### 4.4.1 数据引用

数据引用旨在建立数据与数据之间以及数据与文献之间的关联，进而促进数据的广泛交互。首先 5 个平台中有 4 个平台提供了明确的数据引用方式，其中部分平台只要求用户在遵守相应服务条款的前提下使用或引用数据，如北京大学和复旦大学要求遵循 DataVerse 的引用标准，同时在引用数据时要求标注出处。除去共有的 GB/T 7714-2015 格式，4 个平台提供的数据引用方式有所差别，如北京大学提供 EndNote 格式、RIS 格式和 BibTeX 格式，复旦大学提供 EndNote 格式和 RIS 格式；但浙江大学和华东师范大学仅提供 GB/T 7714-2015 格式；CnOpnData 没有提供明确的数据引用方式,也并未要求用户在引用数据时标记出处和遵循服务条款。

### 4.4.2 数据关联

北京大学、复旦大学、和浙江大学平台所提供的(元) 数据关联的内容较为

单一，均只以超链接的方式关联了数据集的相关数据；华东师范大学平台提供由数据集创造者创建关联导航，可以把其他用户创建的专题链接到本地创建的专题导航上，看起来就像本地创建的专题一样，但点击专题是跳转至其他账户创建的页面上浏览相关专题内容；CnOpenData 平台则将数据集按类型进行分类归纳，因此在一个数据集中除了关联至本身数据集的相关数据，还关联了同一类型的相关外部数据集，另外部分数据集还关联了相关文献。

### 4.4.3 文件格式

文件格式会影响当前和未来软件“导入”数据集的能力，进而影响数据集的解释和理解。

首先是数据提供者上传数据文件格式方面，北京大学平台虽有在使用手册中提及上传文件的信息，但并未提及详细规定的文件格式，同时普通用户在未加入数据集用户组之前并不能上传数据；复旦大学并未规定上传文件的格式，支持不同格式的文件；华东师范大学在使用手册中规定了上传文件的格式，其中包括 Tabular Data、Network Data 和 Big Data 三大类，其中 Tabular Data 分为 SPSS 文件（.sav or .por）、STATA 文件(.dta)、RData 文件，Network 文件就是 GraphML 文件。SPSS）、STAT）、CSV、TAB、GraphML 文件将处理为可设置子集的数据文件，它们可以用数据资源平台分析工具进行在线分析。当选择 CSV（Character-separated Values）数据类型时，首先需要 SPSS 控制卡文件。当选择 TAB（Tab-delimited）数据类型时，首先需要 DDI 控制卡文件。BigData 类型的文件涉及到大数据引擎，需要结合大数据引擎平台进行设置；浙江大学并未提及上传文件的格式； CnOpenDataData 没有上传文件的功能，因此也没有上传文件格式的规定。

在数据使用者下载数据文件格式方面，从八个平台的数据集可以看出，只有小部分数据集提供平台要求上传时的特殊格式，大部分数据集以一些常见的格式呈现，比如纯文本（txt）、数据表（csv、xlsx）、文本文档（pdf、docx）、图片（jpg、png）和程序文件 json 等。因为每个平台的每个数据集提供的数据集数据文件格式都不是固定的，因此对于某个平台存在多种格式的情况，采取“就高不就低”的原则，即将开放度最高的数据格式级别作为整个平台的评分对象。

## 4.5 可重用性

可重用是 FAIR 原则的目标，为了实现这一目标，需要充分描述数据，并在重用过程中明确知识产权，确保数据的可重用性。可重用性的各指标评估结果如表 4-4 所示。

表 4-4 可重用性维度各指标评估结果

Table 4-4 Evaluation results of reusable dimension indicators

| 平台 | | 北京大学 | 复旦大学 | 华东师范大学 | 浙江大学 | CnOpenData |
|---|---|---|---|---|---|---|
| 数据许可 | | | | | | |
| 许可声明明确性 | 许可声明是否明确 | √ | √ | √ | √ | × |
| 许可声明标准性 | 许可声明标准 | CC0 | CC0 | 自定义许可 | - | - |
| | 许可类型 | 遵循标准的机器可读许可 | 遵循标准的机器可读许可 | 未遵循标准的文本许可 | 无重用许可声明 | 无重用许可声明 |
| 限制声明明确性 | 是否存在访问受限数据 | √ | √ | √ | √ | √ |
| | 限制声明与否 | √ | √ | √ | √ | √ |
| | 限制程度 | 有限制重用的数据集，说明限制原因或期限 | 有限制重用的数据集，说明限制原因或期限 | 有限制重用的数据集，说明限制原因或期限 | 有限制重用的数据集，说明限制原因或期限 | 有限制重用的数据集，说明限制原因或期限 |
| 数据溯源 | | | | | | |
| 是否有数据溯源信息 | | √ | √ | √ | √ | × |
| 数据溯源信息标准性 | | 机器可读溯源格式 | 文本溯源格式 | 文本溯源格式 | 文本溯源格式 | 没有描述溯源信息 |

### 4.5.1 许可

如果数据使用规定不够明确，将会限制组织和个人对数据进行再利用。

在许可声明明确性方面，5个平台中有4个平台声明数据使用许可协议，分别是北京大学、复旦大学、华东师范大学和浙江大学，CnOpenData平台未作数据重用规定。

在许可声明标准性方面，有 2 个平台采用遵循标准的机器可读许可标准，仅基于 DataVerse 的北京大学和复旦大学遵循 CC0 协议，即"公共领域贡献宣告"; 2 个平台未使用标准许可协议，华东师范大学使用自己定义的数据使用协议，且内容较为全面且详细，浙江大学则仅有版权及所有权声明；CnOpenDataData 平台未声明。

在限制声明明确性方面，5 个平台均声明未开放的数据集及其受限情形并说明了限制原因，其中北京大学、复旦大学、华东师范大学和浙江大学的数据集限制原因都是数据提交者对数据集设置而导致的，而具体权限分布则在上文可访问性均有提交； CnOpenData 由于并未有上传数据的选项，因此其数据受限的原因是平台会对个性化的数据定制服务数据以及部分数据量庞大以及较高获取难度的数据采取收费行为，需付费成 VIP 才可使用。

### 4.5.2 数据溯源

数据溯源为数据质量的评估提供了解决思路，数据溯源信息主要来自于数据上传者所发布的元数据。数据溯源信息一般包括数据发布和更新时间、数据提交者和联系信息，以及数据集访问地址、出处、版本、元数据标准等。5 个平台中有 4 个平台有较为完善的数据溯源信息，其中做得最好的平台是北京大学，均采用 UNF 数字指纹管理数据集和文件版本，支持历史数据引证，使用标准的机器可读溯源格式。除此之外，另外 3 个平台虽有较为全面的数据溯源信息，提供历史版本的比较等功能，但均未声明其使用标准且均只通过网页文本描述溯源，可见机器可读格式的溯源标准未得到充分应用， CnOpenDataData 则没有数据溯源信息。

## 4.6 研究结论

数据的发现是数据利用的前提。从 Johnson Masinde 的研究[34]中发现高校图书馆数据开放平台通过规范数据标识符的使用、遵循相应的元数据标准和优化数据的检索来支持数据的发现，这一点在国内平台中也得到印证。但相较于国外平台，国内平台仅部分使用国际规范元数据标准，与国际成熟的元数据标准存在一

定差距，不利于数据的长期有效访问，同时各平台间的元数据标准也不统一且不完善；国外平台广泛注册于国际认可的通用仓储，扩大了平台的影响力，易于用户发现数据，国内平台在此方面有待加强。

数据的可访问程度决定了数据可利用的范围。总体而言，用户可以浏览并访问免费开放的图书馆数据开放平台中的大部分数据；调研的大部分平台只提供HTTP 协议访问，在访问协议上缺乏多样性，给用户提供的数据入口数量也很少；在数据的下载获取方面，国内平台都设置了相关授权机制来赋予用户相应的访问权利，有部分平台可以直接下载数据，而有些平台需要注册后才能下载数据，某种程度上给用户带来了不便；在数据的存储方案上，多数平台并未公开说明数据存储机制，数据的长期稳定保存需要进一步的保障。

可互操作涉及语义与技术层面的数据系统建设理念。在数据引用方面，大部分样本平台都有规范且多样的引用方式，但在规定数据引用时的使用条款方面还有所欠缺；在数据的标准化表达方面，国内平台尚未采用语义化语言来表示数据，在平台互操作这一方面还有待提升；在技术层面，互操作意味着不同系统间数据资源的互联，因此，在数据资源中嵌入相关资源的指向链接是实现系统间数据交互的渠道，在这一方面，国内平台的建设普遍优于国外平台。

可重用能够实现数据的反复利用，降低科学研究的成本，实现数据价值的最大化。明确的数据使用许可是保障数据可重用的关键，在这一方面，国外平台已经颁布了官方的数据开放使用许可标准，明晰了数据共享方式、范围和责权，故国外平台普遍使用通用的机器可读许可声明，而国内目前仅少数基于外国数据存储库建立的平台在使用机器可读的许可声明；国外平台已采用溯源元数据标准并纳入元数据，而国内平台在溯源元数据的提供与建设上仍处于起步阶段。

# 5 基于 FAIR 原则的我国高校图书馆数据开平台的改进建议 Suggestion for China academic libraries' open data repositories based on FAIR principle

## 5.1 明确自身定位，发挥核心优势

高校图书馆数据开放平台的核心优势在于高质量的科研数据资源。科研开放数据不仅是一种重要的数据资料，也是促进社会生产力发展的主要推动力。首先，高校图书馆应充分利用其丰富的数据资源，将数据资源有效组织，确保科研数据能够被发现利用；其次，高校图书馆应与其他组织合作，通过开放合作，搭建科研数据与产业需求之间的桥梁，促进跨领域、跨学科的知识和技术交流，从而实现科研数据从理论探索向实际应用的关键转化。通过整合、加工和优化这些数据资源，高校图书馆将有力推动科研数据成为促进科技创新和产业升级的核心生产要素，助推新质生产力的形成发展。

## 5.2 提高数据质量，推动生产要素组合配置

新质生产力是以人工智能、大数据分析等为重要标志的技术创新，其主要驱动力是促进不同产业之间的互联互通和交叉创新，从而形成新的业态和模式[19]。其中数据要素等新兴生产要素与传统生产要素的结合赋能能够提高全要素生产效率。高校图书馆应该加强数据质量监管力度，建立数据质量管理制度，确保数据质量可量化。此外，高校图书馆要使用国际通用的永久性标识符如 DOI、PID、HDL 等替代本地标识符，提高数据的唯一性以及规范性。同时国内数据开放平台还应在多个外部可搜索资源索引处注册，与众多数据仓储库通过数据条目建立链接，以便人和机器探索数据网，即以关联数据的发布方式提高数据的可发现性。最终以高质量科研数据要素做基地，融入到各生产要素中，优化生产过程和效能。

## 5.3 互通平台数据，实现链网联动

我国不同地区开放数据平台的数据存在字段数量、名称、类型等“异构”问题，元数据差异较大，数据互操作性不强[20]。关键的解决办法是确保不同高校图书馆数据开放平台之间的互操作性。各平台应复用已有国际通用的元数据标准，对数据格式进行调整与协调，以机器可理解的格式发布数据，促进数据的自动搜索，增强异构数据的互操作性，以此提高数据在不同平台异构系统的互通性。同时，在数据关联方面，通过链接到数据集的历史版本、相关数据集或资源（如出版物、存储库、平台等），可以提高同一学科、同一地区的数据互通性，以此打破数据孤岛。进而在此基础上，促进科研数据的国际共享。

## 5.4 规范数据标准，优化用户访问体验

根据评估结果，国内高校图书馆数据开放平台在可访问性这一方面存在诸多欠缺。保证用户访问体验可以吸引不同用户人群访问数据，挖掘数据价值，实现数据多样化利用。因而国内高校图书馆数据开放平台需适当开放数据访问权限，可以采取用户分级政策，针对不同类型的用户设置不同的访问权限。同时还需为用户提供多个访问途径，以 HTTP 通信协议为基础，配合使用 URI 来定位与访问资源，使用通用 API 接口作为基本工具开放获取部分数据与元数据，避免出现单一访问途径。此外，平台应声明元数据的长期存储机制和方案，确保数据在可靠的存储库中进行稳定保存和安全传输，保证数据的访问不受时间限制，避免因意外丢失而造成数据无法访问等情况而导致用户使用体验下降。

## 5.5 转化数据效能，促进技术创新

高校图书馆数据开放平台在统一数据标准、互通数据资源的基础上，可以进一步创新数据服务模式，例如为不同专业的科研人员定制个性化服务，为其提供专而精的科研数据。其次高校图书馆也可以开展产学研合作交流，与企业、政府等机构合作，将数据资源与生产需求相对接，促进技术创新和产业发展；支持创新创业项目，利用数据开放平台提供的技术和资源，帮助孵化新技术和新产品。最后也可以建立激励机制，鼓励用户利用开放平台进行科研创作，为数据贡献者

和创新人员提供奖励和资助，以鼓励创新活动蓬勃发展。通过采取多样化手段，促进科研数据的充分利用并转化为实际效能，不断推动科技创新。

## 6 结语与展望 Conclusion and Prospect

本文借鉴国外 FAIR 原则应用评估指标构建了面向我国高校图书馆数据开放平台的 FAIR 应用评估框架。基于对我国 5 个样本平台 FAIR 应用情况的调研分析，从 F、A、I、R 4 个维度分别提出了 FAIR 原则的应用建议。本研究限于调研样本平台数量和调研时间，调研结果对反映我国高校图书馆数据开放平台 FAIR 应用的整体情况存在一定局限。因此，本研究希望通过构建的高校图书馆数据开放平台的 FAIR 应用评估框架及基于调研结果提出的 FAIR 应用建议能为对我国推进高校图书馆开放数据管理、特别是 FAIR 原则的具体应用做出贡献，进一步提高高校图书馆开放数据的开放重用水平，推动我国高校图书馆的数据开放事业发展。

# 参考文献


[1] 推动新质生产力加快发展 [EB/OL]. 2024-02-20[http://www.qstheory.cn/qshyjx/2024-02/20/c_1130079069.htm.
[2] 中共中央 国务院关于构建更加完善的要素市场化配置体制机制的意见 [M/OL]. 2020-05-21[http://www.mofcom.gov.cn/article/b/g/202005/20200502967296.shtml.
[3] 释放数据潜能筑基新质生产力 [EB/OL]. 2024-02-29[https://www.cac.gov.cn/2024-02/29/c_1710883024536400.htm.
[4] 张俊, 谢冰冰. 开放科学环境下的高校图书馆科研角色与服务 [J]. 图书馆论坛, 2021, 41(07): 96-107.
[5] 王永友, 曹兰. 高校要自觉肩负起新时代的文化使命 [J]. 中国高等教育, 2023, (21): 12-5.
[6] 刘萍. 高校图书馆古籍保护育人功能实现路径 [J]. 图书馆工作与研究, 2023, (09): 89-93.
[7] 赵安明, 余超, 张在琛. “双一流”高校建设的责任使命和路径研究 [J]. 东南大学学报(哲学社会科学版), 2023, 25(S2): 63-6+71.
[8] 杨咏梅. 科研数据开放驱动下高校图书馆学科服务转型研究 [J]. 图书馆工作与研究, 2019, (03): 73-7.
[9] 李洋, 温亮明. 我国高校图书馆科学数据开发现状调研与分析——以一流大学建设高校图书馆为例 [J]. 图书馆工作与研究, 2021, (12): 5-15.
[10] 曹筱涵. 开放数据环境下高校图书馆科研数据服务模式构建 [D], 2022.
[11] 陈朝兵, 张田. 2011-2020 年国内外政府数据开放平台建设研究述评 [J]. 图

书情报知识, 2022, 39(01): 130-40.
[12]WILKINSON M D, DUMONTIER M, AALBERSBERG I J, et al. The FAIR Guiding Principles for scientific data management and stewardship [J]. Scientific Data, 2016, 3.
[13]FAIR Data in Earth, Space, and Environmental Sciences [J]. 2018-11-07.
[14]JACOBSEN A, DE MIRANDA AZEVEDO R, JUTY N, et al. FAIR Principles: Interpretations and Implementation Considerations [J]. Data Intelligence, 2020, 2(1-2): 10-29.
[15]Hodson S.FAIR Data:principles,implementation,implications [J]. 2020-03-19.
[16]Zenodo.FAIR Data Action Plan:Interim Recommendations and Actions from the European Commission Expert Group on FAIR Data [J]. 2020-03-19.
[17]Building the data landscape of the future: FAIR Semantics and FAIR Repositories [J].
[18]ELIXIR.Interoperability Platform [EB/OL]. 2020-09-22[https://elixir-europe.org/platforms/interoperability.
[19]WITTENBURG P, DE JONG F, VAN UYTVANCK D, et al. State of FAIRness in ESFRI Projects [J]. Data Intelligence, 2020, 2(1-2): 230-7.
[20]QUAKULINSKI L, KOUMPIS A, BEYAN O D. The Library of Babel as a digital library of human mistakes [J]. 2023 ACM/IEEE Joint Conference on Digital Libraries (JCDL), 2023: 288-90.
[21]WILKINSON M D, SANSONE S A, SCHULTES E, et al. Comment: A design framework and exemplar metrics for FAIRness [J]. Scientific Data, 2018, 5.
[22]李春秋，杜博雅，耿骞, et al. 医学科学数据开放平台 FAIR 原则的应用评估与调查分析 [J]. 图书情报工作, 2022, 66(03): 72-82.
[23]Toward an Implementable Framework of the Fair Data Principles for Earth Science Data Management and Stewardship [EB/OL]. 2020[https://dl.acm.org/doi/10.5555/AAI28155388.
[24]KOSTER L, WOUTERSEN-WINDHOUWER S. FAIR Principles for Library, Archive and Museum Collections: A proposal for standards for reusable collections [J]. Code4Lib Journal, 2018, 40.
[25] GARDIAN FAIR metrics [EB/OL]. 2021-09-20[https://gardian.bigdata.cgiar.org/files/GARDIAN_ FAIR_metrics_guide.pdf.
[26]The Australian research data commons [EB/OL]. 2021-09-20[https://ardc.edu.au/resources/working-with-data/fair-data/.
[27]The CSIRO 5-star data rating tool [EB/OL]. 2021-11-18[http://5stardata.csiro.au/.
[28]FAIRsFAIR Data Object Assessment Metrics: Request for comments [EB/OL]. 2024-04-18[https://www.fairsfair.eu/fairsfair-data-object-assessment-metrics-request-comments.
[29]JACOBSEN A, KALIYAPERUMAL R, BONINO DA SILVA SANTOS L O, et al. A Generic Workflow for the Data FAIRification Process [J]. Data Intelligence, 2020, 2: 56-65.
[30]GOBLE C, COHEN-BOULAKIA S, SOILAND-REYES S, et al. FAIR Computational Workflows [J]. Data Intelligence, 2020, 2(1-2): 108-21.

[31] DEKKER R. Social Data: CESSDA Best Practices [J]. Data Intelligence, 2020, 2(1-2): 220-9.
[32] WITTENBURG P, LAUTENSCHLAGER M, THIEMANN H, et al. FAIR Practices in Europe [J]. Data Intelligence, 2020, 2(1-2): 257-63.
[33] DEN – DE BASIS [EB/OL]. 2024-02-23[https://www.den.nl/.
[34] MASINDE J, CHEN J, WAMBIRI D, et al. Research Librarians' Experiences of Research Data Management Activities at an Academic Library in a Developing Country [J]. Data and Information Management, 2021, 5(4): 412-24.

# Application Evaluation and Analysis of the FAIR Principle in Open Data Repositories of Academic Libraries in China

Zhang Sufang Liang Rui He Guoyu

South China Normal University Guangzhou 510006

**Abstract:** **[Purpose/Significance]** FAIR principle is a data management principle that guarantees the identification, traceability, sharing and reuse of scientific data, and it is of great significance for the improvement of the academic repositories quality by evaluating the application of the FAIR principle in the open data repositories for academic libraries in China. **[Methods/Process]** This study constructs a FAIR assessment framework applicable to China's open data repositories of academic libraries based on the existing FAIR principle framework through network research, and evaluates and analyzes China's academic libraries' open data repositories in terms of the four dimensions, namely, finable, accessible, interoperable, and reusable. [Results/Conclusions] Based on the above research, this study puts forward suggestions for improving the construction of academic libraries' open data repositories from the aspects of clarifying the repositories' own positioning, improving data quality, interoperability of platform data, standardizing data standards, and transforming data effectiveness.

**Key words:** Academic libraries, Open data repositories, FAIR principle

* This work is supported by Chinese Society of Library Science titled "Research on Intelligent Construction of Small and Medium-sized Libraries Based on Knowledge Ecological Chain" (Project No. : 2022LSCKYXM-FZZ-YB001); by Department of Education of Guangdong Province titled "Influencing Factors and Promoting Mechanisms of Knowledge Transfer among researchers in the Guangdong-Hong Kong-Macao Greater Bay Area from the Perspective of Psychological Contract" (Project No. : 2020GXJK171), 2020 Research Project of the 13th Five-Year Plan for Education Science in Guangdong Province.

Authors: Zhang Sufang, PhD, Associate Professor, Email:342208681@qq.com; Liang Rui, Bachelor; He Guoyu, Master student